Diese Website verwendet Cookies von Google, um Dienste anzubieten und den Traffic zu analysieren. Deine IP-Adresse und dein User-Agent werden zusammen mit Messwerten zur Leistung und Sicherheit für Google freigegeben. So können Nutzungsstatistiken generiert, Missbrauchsfälle erkannt und behoben und die Qualität des Dienstes gewährleistet werden.

WEITERE INFORMATIONEN OK

An unending variety of topics

Monday, January 17, 2011

"The Singularity"

A number of my friends have gotten wrapped up in a movement dubbed "The Singularity." We now have a "Singularity Institute", a NASA-sponsored "Singularity University", and so on as leading futurist organizations. I tend to let my disagreements on these kinds of esoteric futurist visions slide, but I've been thinking about my agreements and disagreements with this movement for long enough that I shall now share them.

One of the basic ideas behind this movement, that computers can help make themselves smarter, and that growth that for a time looks exponential, or even super-exponential in some dimensions and may end up much faster than today may result, is by no means off the wall. Indeed computers have been helping improve future versions of themselves at least since the first compiler and circuit design software was invented. But "the Singularity" itself is an incoherent and, as the capitalization suggests, basically a religious idea. As well as a nifty concept for marketing AI research to investors who like very high risk and reward bets.

The "for a time" bit is crucial. There is as Feynman said "plenty of room at the bottom" but it is by no means infinite given actually demonstrated physics. That means all growth curves that look exponential or more in the short run turn over and become S-curves or similar in the long run, unless we discover physics that we do not now know, as information and data processing under physics as we know it are limited by the number of particles we have access to, and that in turn can only increase in the long run by at most a cubic polynomial (and probably much less than that, since space is mostly empty).

Rodney Brooks thus calls the Singularity "a period" rather than a single point in time, but if so then why call it a singularity?

As for "the Singularity" as a point past which we cannot predict, the stock market is by this definition an ongoing, rolling singularity, as are most aspects of the weather, and many quantum events, and many other aspects of our world and society. And futurists are notoriously bad at predicting the future anyway, so just what is supposed to be novel about an unpredictable future?

The Singularitarian notion of an all-encompassing or "general" intelligence flies in the face of how our modern economy, with its extreme specialization, works. We have been implementing human intelligence in computers little bits and pieces at a time, and this has been going on for centuries. First arithmetic (first with mechanical calculators), then bitwise Boolean logic (from the early parts of the 20th century with vacuum tubes), then accounting formulae and linear algebra (big mainframes of the 1950s and 60s), typesetting (Xerox PARC, Apple, Adobe, etc.), etc. etc. have each gone through their own periods of exponential and even super-exponential growth. But it's these particular operations, not intelligence in general, that exhibits such growth.

Pages

- Home

About Me

Nick Szabo

"A premier thinker about history, law and economics, and the lessons they have for security." -- Adam Shostack, Emergent Chaos

"Szabo comes out with these essays that leave me in awe." -- Brian Dunbar

"Reading material that is eclectic, challenging and endlessly fascinating." -- Sean McGrath, Propylon

"Like most blogs worth my attention, this blog is updated only infrequently. That is because the authors of blogs worth my attention only post when they have something to say that is true, relevant and not already known by their audience. Most of the human race does not have the skill to know when an idea has these three properties. The skill is particularly rare in the fields of politics and economics, which is why this blog is such a rare and valuable thing." -- Richard Hollerith

View my complete profile

At the start of the 20th century, doing arithmetic in one's head was one of the main signs of intelligence. Today machines do quadrillions of additions and subtractions for each one done in a human brain, and this rarely bothers or even occurs to us. And the same extreme division of labor that gives us modern technology also means that AI has and will take the form of these hyper-idiot, hyper-savant, and hyper-specialized machine capabilities. Even if there was such a thing as a "general intelligence" the specialized machines would soundly beat it in the marketplace. It would be very far from a close contest.

Another way to look at the limits of this hypothetical general AI is to look at the limits of machine learning. I've worked extensively with evolutionary algorithms and other machine learning techniques. These are very promising but are also extremely limited without accurate and complete simulations of an environment in which to learn. So for example in evolutionary techniques the "fitness function" involves, critically, a simulation of electric circuits (if evolving electric circuits), of some mechanical physics (if evolving simple mechanical devices or discovering mechanical laws), and so on.

These techniques only can learn things about the real world to the extent such simulations accurately simulate the real world, but except for extremely simple situations (e.g. rediscovering the formulae for Kepler's laws based on orbital data, which a modern computer with the appropriate learning algorithm can now do in seconds) the simulations are usually very woefully incomplete, rendering the results usually useless. For example John Koza after about 20 years of working on genetic programming has discovered about that many useful inventions with it, largely involving easily simulable aspects of electronic circuits. And "meta GP", genetic programming that is supposed to evolve its own GP-implementing code, is useless because we can't simulate future runs of GP without actually running them. So these evolutionary techniques, and other machine learning techniques, are often interesting and useful, but the severely limited ability of computers to simulate most real-world phenomena means that no runaway is in store, just potentially much more incremental improvements which will be much greater in simulable arenas and much smaller in others, and will slowly improve as the accuracy and completeness of our simulations slowly improves.

The other problem with tapping into computer intelligence -- and there is indeed after a century of computers quite a bit of very useful but very alien intelligence there to tap into -- is the problem of getting information from human minds to computers and vice versa. Despite all the sensory inputs we can attach to computers these days, and vast stores of human knowledge like Wikipedia that one can feed to them, almost all such data is to a computer nearly meaningless. Think Helen Keller but with most of her sense of touch removed on top of all her other tragedies. Similarly humans have an extremely tough time deciphering the output of most software unless it is extremely massaged. We humans have huge informational bottlenecks between each other, but these hardly compare to the bottlenecks between ourselves and the hyperidiot/hypersavant aliens in our midst, our computers. As a result he vast majority of programmer time is spent working on user interfaces and munging data rather than on the internal workings of programs.

Nor does the human mind, as flexible as it is, exhibit much in the way of some universal general intelligence. Many machines and many other creatures are capable of sensory, information-processing, and output feats that the human mind is quite incapable of. So even if we in some sense had a full understanding of the human mind (and it is information theoretically impossible for one human mind to fully understand even one other human mind), or could somehow faithfully "upload" a human mind to a computer (another entirely conjectural operation, which may require infeasible simulations of chemistry), we would still not have "general" intelligence, again if such a thing even exists.

That's not to say that many of the wide variety of techniques that go under the rubric "AI" are not or will not be highly useful, and may even lead to accelerated economic growth as computers help make themselves smarter. But these will turn into S-curves as they approach physical limits and the idea that this growth or these many and varied intelligences are in any nontrivial way "singular" is very wrong.

Posted by Nick Szabo at 9:32 PM

Blog Archive

- ► 2018 (1)
- ► 2017 (3)
- ► 2016 (4)
- ► 2015 (3)
- ► 2014 (3)
- ► 2013 (3)
- ► 2012 (8)
- ▼ 2011 (12)
  - ► December (1)
  - ► July (1)
  - ► June (2)
  - ► May (2)
  - ► February (2)
  - ▼ January (4)
    - Some hard things are easy if explained well
    - Tech roundup 01/22/11
    - "The Singularity"
    - Making a toaster the hard way
- ► 2010 (9)
- ► 2009 (29)
- ► 2008 (55)
- ► 2007 (47)
- ► 2006 (130)
- ► 2005 (44)

68 comments:

Anonymous said...

The singularity refers to the time in human history when computers first pass the Turing test. There is nothing "incoherent" or "intrinsically religious" about this idea. It's obviously going to happen someday; the only question is when.

Yes, "all growth curves that look exponential or more in the short run turn over and become S-curves or similar in the long run". But one can argue about the period of time to which "the long run" refers in a particular case. For instance, the number of humans and their descendants in the universe may continue to grow exponentially until begin to expand outward in a sphere at the speed of light, at which point the growth is obviously no longer exponential.

Obviously we can't predict the price of stocks or the weather next week, but we have an idea of what the world will be like in general. But the singularity is a bigger change than a stock market crash. It seems to me that we can't predict the parameters of a post-singularity world any better than a prelingual human could have predicted what human life would be like after the invention and widespread adoption of language.

What you call the “singulatarian notion of an all-encompassing or "general" intelligence” is due to Turing, and the existence of such an intelligent nonhuman isn't prohibited by any economic principle. In fact, the market provides an incentive for the creation of these nonhuman intelligences. For instance, one of Google's goals is to be able to answer any question a user asks. They're a long way from achieving this goal, but there should be no doubt that this service would be in demand, even in a marketplace filled with specialized non-general intelligences.

Finally, it may indeed turn out to be infeasible to upload a human brain into a computer, but that's just one of many ways the singularity could come about.

10:50 PM

Rupert Goodwins said...

There's also the small question of what intelligence actually is, and how it may be bounded. That's rarely discussed in this context - after all, when machines get smart enough they'll demonstrate the answer, right?

I think this is a real problem, and an interesting one. If you think of intelligence as the ability to create models based on data and come to correct conclusions about hidden information (either hidden because it hasn't happened yet, or because it has but you haven't experienced it), then the limit comes at the speed you can absorb and act on data. That must be bounded by c.

As even today's meatspace engineers know, the limits on computation quickly converge on the limits on data transfer. We can build 1000 core CPUS and beyond (back of the envelope calculations say that you could put something like a third of a million 1980-vintage 8-bit processors on one current high-end production die): that's not that useful. All the effort goes into making sure that the cores can get to and share data, and they can't do that very well.

Whatever intelligence is, it won't be able to scale beyond the limits where you just can't move data or internal computations around within the mechanism fast enough for the extra smarts to be able to that useful.

That cosmic gestalt mind? Nah.

11:13 PM

nick said...

The singularity refers to the time in human history when computers first pass the Turing test.

That's one definition among many. And which Turing Test? There are an infinite number of possible Turing tests, some of them easy, some of them hard. e.g. The "captcha" test we go through to register for web sites, some versions of which some computers can already pass. If the visual-letter captcha becomes obsolete, it doesn't herald The Singularity, although it may play some hob with computer security as we scramble to come up with a less spoofable (probably because higher bandwidth, e.g. sound) variation on captcha.

but we have an idea of what the world will be like in general.

Really? Twenty years ago were you envisioning Facebook and the popularity of texting among teens? What percentage of people in the early 1950s envisioned life two decades later after the sexual revolution? How many people in the 1440s envisioned life in a society where most people were literate? Or as you point out, the change from prelingual to lingual humans. Unexpected qualitative changes have been happening for a very long time. They weren't "singularities" in any meaningful sense of that term, they were qualitative, complex changes that evolved over time. They were unpredictably unpredictable, not predictably unpredictable.

11:46 PM

5678 said...

It seems that arguing against the singularity without having compelling evidence to support the hypothesis should too come under this religious dub, as it is holding a belief without vital evidence. The null hypothesis would be: there does not exist an algorithm and a computer which achieves the goals of AGI. I await the proof, or a modicum of scientific evidence to support this.

11:53 PM

Anonymous said...

And which Turing Test?

The Turing Test refers to the imitation game introduced in Turing's 1950 paper "Computing Machinery and Intelligence". In today's terms, a computer would be said to pass this test if skilled interlocutors are unable to distinguish between the computer and a human over suitably long IM conversations. By the way, there exists (in the mathematical sense) an algorithm which passes the Turing Test.

11:55 PM

nazgulnarsil said...

I don't think it takes a religious belief to agree with one of two propositions: that we will invent a general intelligence capable of understanding and improving upon its own design or that we will emulate a human brain in silicon (or some other substrate) that results in a significant speed up in human intelligence. either of these events will dramatically alter the human future.

12:00 AM

Kevembuangga said...

Wow!
I am afraid you've triggered an avalanche of Singularity bigotry.
The comment from 5678 asking for evidence "against" very much reminds the arguments of fundamentalists v/s atheists.

More to the point, some singularitarians pretend to have a definition of "Universal Intelligence" which curiously enough isn't

computable. :-)

12:20 AM

nick said...

Anonymous commenters, please choose a handle, thanks. -- the management

12:24 AM

Alrenous said...

Wherein I attempt to make a point almost entirely through quotation and juxtaposition.

"We have been implementing human intelligence in computers little bits and pieces at a time [...] But it's these particular operations, [boolean logic, linear algebra, typesetting] not intelligence in general, that exhibits such growth."

vs.

"invent a general intelligence"

and

"information and data processing under physics as we know it are limited by the number of particles we have access to, and that in turn can only increase in the long run by at most a cubic polynomial"

vs.

"emulate a human brain"

http://blog.createdebate.com/2008/04/07/writing-strong-arguments/

For completeness, my position is that the issue is dominated by unknown unknowns and therefore unwarranted confidence. Considering the number of surprising setbacks and redirects even small, well-understood engineering projects encounter, and taking the limit outward toward large and poorly understood...

12:32 AM

Kevembuangga said...

About *the problem of getting information from human minds to computers and vice versa* and *informational bottlenecks* do you know of this approach which tries to at least partially care for the poor performance of human mind with complex rules.

12:33 AM

bumbledraven said...

Anonymous commenters, please choose a handle

Good idea. I'm anonymous #1 here. I'm also also the person who submitted your post to Hacker News.

12:38 AM

5678 said...

@Kevembuangga

The comparison to fundamentalists vs atheists is an invalid and inaccurate one. My argument was simple, mathematical logic. Given a logic statement, establishment of absolute truth is only accepted through rigorous proof. This is the basis of mathematics, and the sciences in general.

Theorem 2: The analogy is inaccurate.

Proof: The theistic debate is between gnostics and agnostics, not agnostic atheists and gnostic theists.

12:48 AM

nick said...

Thanks, bumbledraven. As for the Turing Test, even keeping strictly to the original paper's definition (rather than a common modern definition that includes things like "captcha" where the "interlocutor" is itself a computer) there are a very large number of possible skilled interlocutors, with some ill-defined and thus varying skills levels. Even four decades ago, the chat bot (as we might call it today) "Eliza" was fooling some people into treating it as human. The ability to passably talk like a human, especially over such an abstract medium as text, is not anywhere close to the same as being able to think like a human.

We've gone through a large number of Moore's Law doublings since Eliza and no AGI apocalypse has resulted. And the proportion of humans that can discover the ruse probably hasn't changed all that much, strongly suggesting that this (specialized, like all other) AI skill improves far more slowly than the cost of memory, CPU, etc. falls. Simarily, machine learning capabilities don't double with these hardware improvements because learning curves are typically logarithmic.

1:08 AM

nick said...

5678, perhaps @Kevembuangga was thinking more of medieval Christian theologians than of today's fundamentalists. These scholars, very well credentialed as the first recipients of the "Doctor of Philosophy" (PhD) degrees, were very knowledgeable about logic, and used it to prove, based on a variety of questionable assumptions, the existence of God. And of course, they demanded that their opponents rigorously *disprove* the existence of God, which is about the same thing as asking for a mathematical disproof of The Singularity. Very rigorous fellows, they were. Albeit not very scientific, IMHO.

1:21 AM

Jinzang said...

"medieval Christian theologians ... demanded that their opponents rigorously disprove the existence of God"

Which version of history are you reading from? Because it doesn't resemble any I'm familiar with. Who are these medieval atheist philosophers?

6:12 AM

gwern said...

> The "for a time" bit is crucial. There is as Feynman said "plenty of room at the bottom" but it is by no means infinite given actually demonstrated physics. That means all growth curves that look exponential or more in the short run turn over and become S-curves or similar in the long run, unless we discover physics that we do not now know, as information and data processing under physics as we know it are limited by the number of particles we have access to, and that in turn can only increase in the long run by at most a cubic polynomial (and probably much less than that, since space is mostly empty).

Yes, but the fundamental limit is so ridiculously high that it might as well be infinite. Look at Seth Lloyd's bounds in http://arxiv.org/abs/quant-ph/9908043 . He can be off by many orders of magnitude and still Moore's law will have many doublings to go.

(Incidentally, the only quasi-Singulitarian I am aware of who has claimed there will be an actual completed infinity is Frank Tipler, and as I understand it, his model required certain parameters like a closed universe which have since been shown to not be the case.)

> As for "the Singularity" as a point past which we cannot predict, the stock market is by this definition an ongoing, rolling singularity, as are most aspects of the weather, and many quantum events, and many other aspects of our world and society. And futurists are notoriously bad at predicting the future anyway, so just what is supposed to be novel about an unpredictable future?

WHAT. We have plenty of models of all of those events. The stock market has many predictable features (long-term appreciation of x% a year and fat-tailed volatility for example), and weather has even more (notice we're debating the long-term effects of global warming in the range of a few degrees like 0-5, and not, I dunno, 0-1,000,000). Our models are much better than the stupidest possible max-entropy guess. We can predict a hell of a lot.

The point of Vinge's singularity is that we can't predict past the spike. Will there be humans afterwards? Will there be anything? Will world economic growth rates and population shoot upwards as in Hanson's upload model? Will there be a singleton? Will it simply be a bust and humanity go on much as it always has except with really neat cellphones? Max-ent is our best guess; if we want to do any better, then we need to actively intervene and make our prediction more likely to be right.

6:58 AM

nazgulnarsil said...

I'm sorry alrenous but in this context we're obviously reduced to pointing at strong arguments and not making them ourselves.

to assign very low probability that either proposition I mentioned comes true by 2100 is fine. you are well within reasonable estimations: http://theuncertainfuture.com/index.php

7:18 AM

bumbledraven said...

There are a very large number of possible skilled interlocutors, with some ill-defined and thus varying skills levels.

To fix a point of reference, let us consider the version of the Turing Test specified in the terms of the $20,000 wager between Kapor and Kurzweil ("By 2029 no computer - or "machine intelligence" - will have passed the Turing Test.").

The ability to passably talk like a human, especially over such an abstract medium as text, is not anywhere close to the same as being able to think like a human.

Really? If passing the Turing Test is not enough, what would it take to convince you that an entity is indeed able to think like a human?

8:32 AM

bumbledraven said...

Link to terms of the wager between Kapor and Kurzweil: http://www.longbets.org/1

8:39 AM

Emile said...

Aw please, could we avoid dragging the discussion down with analogies to religion and theology and fundamentalists?

About the definition of "Singularity": This article presents three different definitions of "Singularity" that are used ... I don't think many people think of a literal "mathematical singularity" (you seem to be talking about all three approaches instead of just focusing on one, so I'd bet you already read that article).

Good points about unpredictability, the future is pretty hard to predict, singularity or no singularity. It does seem to me, though, that the "time before the future becomes totally baffling" has been getting shorter and shorter (i.e. someone from 1600 teleported to 1650 wouldn't find life surprising a someone from 1960 teleported to 2010). Rudyard Kipling might disagree though.

Nor does the human mind, as flexible as it is, exhibit much in the way of some universal general intelligence. [...] So even if we in some sense had a full understanding of the human mind [...], we would still not have "general" intelligence, again if such a thing even exists.

Human-level intelligence in machines (AI or uploads), whether or not you call it "general intelligence" is enough to be a game changer; once a machine can do pretty much anything an average human can do through a computer (program, interpret statistics, run a business), then the economy is in for some big changes - *everything* changes. Well, a lot of things do at least.

8:57 AM

nick said...

Who are these medieval atheist philosophers?

I imagine you've heard the phrase "devil's advocate"? That's where it comes from. They weren't supposed to win, of course, just make a good argument to exercise the rhetorical and logical skills of their fellow Christian theologians.

10:18 AM

Emile said...

Good points about unpredictability, the future is pretty hard to predict, singularity or no singularity. It does seem to me, though, that the "time before the future becomes totally baffling" has been getting shorter and shorter (i.e. someone from 1600 teleported to 1650 wouldn't find life surprising a someone from 1960 teleported to 2010). Rudyard Kipling might disagree though.

Nor does the human mind, as flexible as it is, exhibit much in the way of some universal general intelligence. [...] So even if we in some sense had a full understanding of the human mind [...], we would still not have "general" intelligence, again if such a thing even exists.

Human-level intelligence in machines (AI or uploads), whether or not you call it "general intelligence" is enough to be a game changer; once a machine can do pretty much anything an average human can do through a computer (program, interpret statistics, run a business), then the economy is in for some big changes - *everything* changes. Well, a lot of things do at least.

(Was the previous version of this comment deleted?)

10:21 AM

nick said...

Hi, my apologies for the commenting problems, they come in two forms:

(1) Posts that contain insults directed at specific individuals get deleted as soon as I discover them.

(2) Blogger is putting many (but puzzling not all) posts that contain links in the "Spam" category, which means these posts are basically being moderated.

For category (1), please take your idiotic bigotry elsewhere, for category (2), I apologize and will try to check the Spam queue more frequently as long as this discussion continues.

10:35 AM

Alrenous said...

Hey nick, why do you care whether the singularity idea is wrong or not? What are you going to do with this conclusion?

Hey nazgulnarsil, why do you care whether...?

Hey anyone else who just feels like answering, why do you...?

Just for my own curiosity, nazgulnarsil, why does it matter to you whether nick thinks the Singularity is Religious(TM) or not?

10:38 AM

nick said...

P.S. to my last comment: Emile, obviously your post was in category (2), Blogger pathologically categorized it as "Spam".

gwern:
the only quasi-Singulitarian I am aware of who has claimed there will be an actual completed infinity is Frank Tipler,

Then Vinge's analogy to the infinite density of a black hole (origin of the term "The Singularity") is a rather bad one isn't it? Why use a bad analogy to label the movement?

Emile:
It does seem to me, though, that the "time before the future becomes totally baffling" has been getting shorter and shorter

Although there have been plenty of "spikes" where the future has become in many ways baffling to prior generations, per the examples I gave above. It is probable in terms of the raw complexity of society,if we could measure such a thing, that we are still either on the first leg of the S-curve inflexion or at least not so far past the inflexion that we notice things slowing down.

Of course, as I stated different pieces follow (to the extent they can be said to follow curves at all) very different S-like curves. Growth has slown down greatly in certainly areas, e.g. the growth in performance of transportation which was so dramatic in the 19th and most of the 20th century but since the 1970s has proceeded at an arduous crawl and in some ways (e.g. the efficiency of public transportation) may have even declined. Even many IT indicators are now growing far more slowly than Moore's Law, for

example CPU speed. We face many practical limits long before we face the ultimate physical limits, and whether or how we may surpass such limits is unpredictable (there's often great truth to the saying that we don't predict the future, we make it).

11:14 AM

nick said...

Hey nick, why do you care whether the singularity idea is wrong or not?

Well I mostly didn't care much about it until last night, when I got a bee in my bonnet. :-) Maybe I won't care about it tomorrow, either, who knows. Predicting what I'm going to care about in the future is my own individual singularity. :-)

Currently, however, I am bugged by certain friends, who shall remain nameless, who insist that conversations be dominated by bad analogies ("singularity"), vague hypotheticals ("general intelligence"), and the like. For a long time I have preferred studying things that have actually happened (e.g. law, actual technology, history) to things that might or might not happen according to vague imaginings. That way my brain gets filled with real information rather than junk and I actually have a better grasp on what might happen in the future than people who don't understand the past. IMHO, your mileage may vary, etc.

11:28 AM

gwern said...

> Then Vinge's analogy to the infinite density of a black hole (origin of the term "The Singularity") is a rather bad one isn't it? Why use a bad analogy to label the movement?

It seems like a perfectly good analogy to me.

> I think it's fair to call this event a singularity ("the Singularity" for the purposes of this paper). It is a point where our old models must be discarded and a new reality rules.

http://mindstalk.net/vinge/vinge-sing.html

> We will soon create intelligences greater than our own. When this happens, human history will have reached a kind of singularity, an intellectual transition as impenetrable as the knotted space-time and the center of a black hole, and the world will pass far beyond our understanding. This singularity, I believe, already haunts a number of science-fiction writers. It makes realistic extrapolation to an interstellar future impossible. To write a story set more than a century hence, one needs a nuclear war in between ... so that the world remains intelligible.

(_Omni_ oped by Vinge)

> Mathematical singularity, a point at which a given mathematical object is not defined or not well-behaved, for example infinite or not differentiable

http://en.wikipedia.org/wiki/Singularity

Hm. Why would he use such a term...

And BTW, black holes are not known to be infinitely dense. What exactly goes on inside a singularity is unclear (as one would expect of something crossing from relativity to quantum mechanics's bailiwick! and as one would expect from all the different speculation & models about naked singularities, white holes, black-hole universes etc.); eg.

> Some theories, such as the theory of Loop quantum gravity suggest that singularities may not exist. The idea is that due to quantum gravity effects, there is a minimum distance beyond which the force of gravity no longer continues to increase as the distance

between the masses become shorter.

http://en.wikipedia.org/wiki/Gravitational_singularity#Interpretation

11:42 AM

Kevembuangga said...

@Alrenous
why do you care whether the singularity idea is wrong or not?

Since I am *truly* interested in AI progress I find the Singularity hype highly detrimental.

1) Paranoid scare about "big bad AI" is hampering AI research funding.

2) Many purported Singularitarians are actually working *against* progress in AI and related fields in spite of their pretense to the contrary.

3) The whole efforts, energy, budgets, PR, etc... devoted to Singularity "watch" could have much more productive uses.

4) Overselling the prospects of AI will bring a backlash as it did already in the 80's.

5) There are many, many *much more realistic* "existential risks" to civilization and mankind than AI overtaking humans.

And, YES, all this stems from religious thinking, a pseudo-secular form of Eschatology.
Stupid monkeys projecting the fear of their own death onto an end-of-the-world scenario.
This shows all over history, the "good old times" are gone and we are about crash, this is somehow often true but not for the reasons invoked.

12:18 PM

Kevembuangga said...

More arguments against some other aspects of the Singularity: The Undesigned Brain is Hard to Copy.

12:44 PM

Jinzang said...

I imagine you've heard the phrase "devil's advocate"? That's where it comes from. They weren't supposed to win, of course, just make a good argument to exercise the rhetorical and logical skills of their fellow Christian theologians.

Your confabulation of medieval history is more entertaining than the real thing. The "devil's advocate" is a priest assigned the task of arguing against the canonization of a would-be saint.

Where exactly is the medieval argument against the existence of God, even if only rhetorical?

3:46 PM

Mentifex said...

The TechnoSingularity ought to arrive sometime before the end of 2012.

4:27 PM

nick said...

Meanwhile Timothy B. Lee makes a good argument about how it will probably always be infeasible to digitally emulate the evolved and analog brain, and the crucial distinction between simulations (which are and may of necessity always be woefully incomplete in this area) and emulation (which must be complete to work). I've clarified and extended his argument as follows:

The most important relevant distinction between the evolved brain and designed computers is abstraction layers. Human engineers need abstraction layers to understand their designs and communicate them to each other. Thus we have clearly defined programming languages, protocol layers, and so on. These are designed, not just to work, but to be understood by at least some of our fellow humans.

Evolution has no needs for understanding, so there is no good reason to expect understandable abstraction layers in the human brain. Signal processing may substantially reduce the degrees of freedom in the brain, but the remaining degrees of freedom are still likely to be astronomically higher than those of any human-understandable abstraction layer. No clean abstraction layers, no emulation.

11:48 PM

Zach K said...

An important point as been missed: an unfriendly singularity has the potential to be so bad that even people who assign a really low probability to any AI-related singularity event should still be concerned about it.

And just how much resources are being investing into singularity-AI work? My guess is not much compared to most other CS research and even AI work, and probably most of that is private money.

The point of SIAI (Sing inst. for AI) is to make sure the problem of friendliness gets solved before we develop AGI: not only do we not know how to define intelligence well, we have no idea how to encode "friendly" motivational structures.

I don't see where the waste of time or resources is: it's not your time or resources.

11:53 PM

nick said...

Zach K, we should be highly suspicious of these Pascal Wager type scenarios: very high payoffs or costs with very low probabilities. One of the problems is that a probability some people guess at 1% could almost as easily be 0.001% or 1-millionth of 1% or even smaller. All of those are well within the margin of error of what we know about "general intelligence", namely extremely little, assuming that it is even a coherent idea. One can make up an infinite number of futuristic scenarios with extreme outcomes not ruled out by the laws of physics. That doesn't make any of them significantly probable, indeed the average probability of these infinite number of disaster scenarios is indistinguishable from zero. Worse, vague claims such as those about AGI are unfalsiable and thus fall outside the world of testable scientific discourse.

In other words, just because one comes up with quasi-plausible catastrophic scenarios does not put the burden of proof on the skeptics to debunk them or else cough up substantial funds to supposedly combat these alleged threats. We would end up spending an infinite amount of money (if we had it) pursuing scenarios that each have an average probability of practically zero. As with more probable scenarios, the burden of proof is on the people asking for money to show why they need it.

When probabilities are so uncertain it is worse than useless, it is false precision, to throw out numerical probability estimates. The focus has to be on the *assumptions* we are making. My article above is an in-depth exploration of many of the assumptions surrounding "The Singularity" and AGI. Both economic considerations (the extreme division of labor that makes up the high-tech

economy and the resulting hypersavant/hyperidiot/hyperspecialized computing systems) and evolutionary/knowledge considerations (the extreme difficulty and indeed probable practical impossibility of emulating the evolved and analog brain on a designed and digital computer) strongly suggest that the AGI/Singularity scenario is highly implausible. The burden of proof is on the catastrophists to show why this standard economics and straightforward distinction between evolved brain and designed computer doesn't apply before their claims can be taken seriously.

Is for SIAI, or any other such organization, they can hardly "make sure" or even make significant positive differences about anything in these extremely uncertain areas. Vague theorizing about alleged hypothetical future "AGI" (which, by some mysterious power, not actually specified as an implementable algorithm, is somehow much more powerful than very general techniques that actually work such as evolutionary algorithms) is probably at best useless and at worst misleading.

1:49 AM

Anonymous said...

Flight. The sound barrier. The Moon.

3:59 AM

nazgulnarsil said...

no particular reason alrenous. I highly respect the author of this blog and thus it pains me to see him disagree seriously with me on a subject. silly monkey-ally-building circuitry firing.

rationality related debate that does not result in "shut up and get rich" is all pretty much just entertainment no?

4:45 AM

Kevembuangga said...

@Zach K
The point of SIAI (Sing inst. for AI) is to make sure the problem of friendliness gets solved before we develop AGI: not only do we not know how to define intelligence well, we have no idea how to encode "friendly" motivational structures.

This is self refuting in two distinct ways:

1) By the very definition of "The Singularity" what comes after is *unpredictable*, even less so controllable.

2) By trying to "monitor" a higher intelligence for friendliness you pretend you will be more clever than it/him/her, isn't that presumptuous to the point of being totally daft?

Even setting aside the implausibility of "The Singularity" the concerns brought up have no logical consistency, a hallmark of *religious thinking*!

5:30 AM

Alrenous said...

nazgulnarsil,

My own interest is epistemological. What can I learn about epistemology by using the singularity question as a test case? I eventually plan to codify a procedure and then test it against actual historical qualitative shifts, to see if it gets reasonable results. Not just entertainment...but as it happens, I do find the process entertaining.

I haven't yet been able to reject my null hypothesis: "I don't know." By all means, point me at some arguments if you don't care to provide them directly.

Kevembuangga,

I have to object to the religious angle.

If someone's interfering with AI research, simply credibly demonstrate to them that they're doing so and how. If they genuinely care, they'll stop. If they genuinely don't, I don't see how calling them religious will make them start caring.

Similarly, if pursuing singularity initiatives fails tests of productiveness, then that means those pursuing them don't test for productiveness before they commit. Even if you manage to stop them wasting stuff on singularity, then only by chance will they land on something useful the next go-around.

nick,

Oddly, I agree with your conclusions even though I can provide two separate concrete definitions of general intelligence.

First is simply Turing completeness. Given enough transistors, there's no logical operation or information manipulation our current machines cannot carry out. The differences between these and the prefrontal cortex are entirely in scale and optimization type. (Re-cast the contents of conscious thought as an extremely high level programming language. Since axons are so slow compared to silicon, perhaps the might of the brain is simply that neurons solve the data transfer issue.)

Second is in terms of information:
Gather information, learning; extrapolate and interpolate that information, reasoning; and generate information, creativity.
I'm pretty sure those three span the entire space of things one can do with information. The general term 'intelligence' is simply the conflation of all three.

To see the scale/optimization differences between the von Neumann machine and the prefrontal cortex, note we already have machines with general intelligence - and look at how stupid they are. (http://singularityhub.com/2009/10/09/music-created-by-learning-computer-getting-better)

7:28 AM

Kevembuangga said...

@AIrenous

If they genuinely don't, etc...

You are likely mistaken about the purpose of my discourse, I certainly do not expect to change the opinion of Singularitarians, I only want to expose their silly ideas for what they are as a warning to other people not to take them seriously and be wary of the waste of resources they represent.
As Nick put it, not to be induced to *"cough up substantial funds to supposedly combat these alleged threats"*.

Incidentally I also disagree with your views on general intelligence:

Given enough transistors, there's no logical operation or information manipulation our current machines cannot carry out. The differences between these and the prefrontal cortex are entirely in scale and optimization type.

And:

Second is in terms of information:
Gather information, learning; extrapolate and interpolate that information, reasoning; and generate information, creativity.

In both cases you are eluding a critical step: It is true that computers or any turing machine equivalent can massage information toward whatever purpose you will but you can only speak about "information" *AFTER* the amorphous raw inputs have been *encoded*, i.e. discretized and labeled with respect to some ontology of objects/concepts/ideas.
This is the critical step which is overlooked in current AI research, why do we *choose* to "recognize" a triangle versus a square in order to model the world?
(or cats or substances or molecules or quarks or verbs or consonants or love or law or whatever...)

It's obvious!
For me too, but why is it obvious?

Taking the objects and concepts as "a given" is the Platonist position, I am anti-Platonist.

If your interest is epistemological I expect you to understand this puzzling question.

9:18 AM

Alrenous said...

Kevembuangga,

"As Nick put it, not to be induced to "cough up substantial funds to supposedly combat these alleged threats"."

I see.

Yes, computers cannot understand the information they process, merely manipulate it. The ultimate decode step must always, as of yet, be performed by a human. As per my music example - it has to be told what humans find pleasing, else it will just serve up random noise. Similarly goals etc.

So, total agreement. Though, this is one of my current research projects. I even have an answer.

The answer sounds completely insane. I'm looking for errors.

9:50 AM

Michael Anissimov said...

Wouldn't more generalist, human-surpassing eventually be developed, even if only specialized AIs were developed at first? And wouldn't those pose a massive potential risk or benefit to humans?

12:28 PM

nick said...

Wouldn't more generalist, human-surpassing eventually be developed, even if only specialized AIs were developed at first? And wouldn't those pose a massive potential risk or benefit to humans?

Michael, thanks for your thoughtful comment. When eventually comes, chances are

(1) an extremely high variety of highly specialized software will already be doing 99.999..(many more places)% of the intellectual heavy lifting already, with humans almost entirely devoted to entertaining each other. An economy of sports, nail salons, and the like, i.e. people basically devoted to providing emotional comfort and stimulation to each other, the intellectual labors having long since been divided to extreme degrees among the quadrillions of kinds of hyperidiot/hypersavant software. The problem of computers replacing humans, requiring humans to retrain themselves on other tasks, comes along gradually, and indeed has been going on since adding machines started slowly replacing mental arithmetic. There is no singular point, no "singularity" at which a big mass humans all of a sudden become obsolete. It's a boiling frog problem. Any general AI will be extremely uncompetitive (whether in the market or in war) with specialty software and will pose no threat to humans that specialty AI hasn't already posed many orders of magnitude greater. All the important threats will come from specialized software.

(2) We still won't be anywhere close to being able to emulate a human brain on a computer (for reasons Tim Lee's linked article and my comment above describe),

(3) Computers by this time, which is still well in the future, will with high likelihood be well up the S-curve of computer performance and it will be growing slowly. In some ways, e.g. CPU speed, growth is already far below exponential, well on the second leg of the S-curve.

For these reasons, and because historical facts as actual facts are far more informative than speculative guesses out of the vast space of possible futures, if we want to study the risks posed by computers, we should study the risks we have already seen to be posed by computers. Not that these risks won't over the years qualitatively change, just that it's pointless trying to predict what those changes will be, and the hopelessly vague and extremely uneconomical idea of "general AI" doesn't shed much light on the issue.

7:00 PM

Kevembuangga said...

@Alrenous
So, total agreement.

May be not.
I don't see anything special in the way humans are chopping down reality into into "significant chunks" with respect to their own interests/motivations.
To me it's only some kind of feature extraction out of raw data akin to what happens in PCA (Principal Components Analysis) which could end up in totally different ontological schemes if we were to have different sensing capabilities (seeing UV, sensing magnetic fields, etc...) or different biological drives (laying eggs instead of babies, eating krill or whatever).
I don't see anything "mysterious" in the notion of *meaning*, only something which is tied to the observing subject, therefore any process elaborate enough to collect and organize a lot of information about its environment could be said to have meanings and purposes of its own.
This is close to the Singularitarians scare but I don't think such an "artificial organism" can be *designed* (re Timothy B. Lee) and it would be utterly cretinous to try to build one.
Self improving AI does not imply that it has to be given full autonomy in its goals, it is just another tool than we can build to manage *our* own goals.

12:14 AM

nick said...

Intriguingly, while Adam Smith et. al. have, as I described in the linked article above, long observed the big wins from specialization in the general economy, one can demonstrate this specifically for the Hutter's AIXI model of "general intelligence" or "universal algorithmic intelligence" of which Singulatarians are fond. Hutter's model is basically just algorithmic probability and Solomonoff

induction which I describe here. As Kevembuangga states this is uncomputable: there is in fact no way to guarantee a computer will find the best answer to any given problem, the best model to describe or predict any given environment, and the like.

Furthermore, even Hutter's time-bounded version of AIXI, which is not guaranteed to find the best answer, requires a time exponential in the size of the analyzed environment to run. By comparison, cracking public-key cryptography takes less time: it's only "super-polynomial" (slower than polynomial but faster than exponential) in the size of the key. Most practical algorithms run in logarithmic time (i.e. doubly exponentially faster than "universal AI"), polynomial time (exponentially faster than "universal AI"), or similar.

And the most practical machine learning techniques, such as evolutionary algorithms, which Singulatarians love to decry as "specialized", since they don't actually produce anything close to human-like intelligence, take advantage of probablism, keeping the best prior answers around, etc. to often get a "good" model of the environment or solution to a problem early in its run. Their performance follows the well-known logarithmic learning curve, i.e. it gets exponentially harder to proceed as one approaches the best answer, so evolutionary algorithms and most other machine learning techniques tend to only work well where "low-hanging fruit", such as readily learned patterns, exist in the environment and are sufficient to solve one's problem (and, as stated above, where environments are easily simulable, which is the hard part).

Thus, in the general AIXI case of Kolmogorov/Solomonoff/Hutter, where the probability distribution of the environment is unknown, there is no guarantee one can find an answer and even where there is no guaranteed answer analyzing the environment requires an infeasibly large amount of time. When one knows some useful things a priori about *particular* environments, i.e. when intelligence is *specialized*, computations that learn things about or solve problems within that environment can often be exponentially or doubly-exponentially more efficient. Thus computation theory proves for intelligence what Smith, Hayek, Reed, et. al. observed about economies: specialization and the distribution of knowledge brings huge advantages.

2:37 AM

Emile said...

nick: I agree that a "S-curve" is fairly likely, and that trying to extrapolate growth curves is kinda silly - the term "Singularity" is maybe useful when talking to lay audiences, or in science fiction, but I think people like Eliezer Yudkozsky, Robin Hanson or Ben Goertzel talk about more precise things (at least, I don't remember Eliezer or Robin spending that much time talking about "the Singularity" in general).

(Damn, my comment was too big, I'll split it up. I hope the spam-eating monster doesn't eat the second half, it has a few links)

6:20 AM

Emile said...

One of Tim's arguments (in a comment to Robin's post) is that software "human-level" AI is easier than brain emulation: *"This isn't to say we couldn't build AI systems to do many of the things human brains do. But we're not going to accomplish that by emulating human brains. We're much more likely to accomplish it by re-implementing those capabilities using the very different techniques of human engineering."* - this is different from your position that both "human level" AI and brain emulation are both unlikely (since either one would be enough to make something like a "Singularity" quite likely).

What probability would you assign to "assuming we don't develop human-level AI or have a major civilization collapse first, we'll be able to run a functional simulation of a human brain (from a scanned brain) in the next 200 years"?

I would consider such a scenario more likely than not. Yes, the brain is a complex system built by natural selection, which doesn't follow the constraints of human engineering, which would make it *easier* to simulate. But there are still *some* structure, especially at the neural level, and much can be learn (is being learned) by studying insect brains, which would be easier to simulate. Drosophila can learn, have "only" 100 000 neurons which would make simulating one significantly easier than a human (see here for work on that), and would help us make sure our model of a neuron is good enough"before we move to something bigger like a frog or a rat. It's

not as if simulating a whole human brain had to be the first step (I agree that would be quite difficult).

Also, we can take some shortcuts natural selection couldn't - maybe some groups of neurons (ganglia or nuclei) can be modeled as a relatively simple system, simpler (and cheaper to emulate) than the sum of the neurons in it. Maybe neurons used for different parts of the brain don't always use all of their functionality and simpler models can sometimes be used (That's the kind of stuff that would be relatively easy to try once we have functional insect brain emulations).

Seems to me the road to human brain emulation has a lot of possible paths; there are some hurdles (computing power, brain-computer interface, scanning resolution) but those are not very mysterious (they don't require fundamental insights the way a lot of AI does) and are in domains which are currently progressing.

Which part exactly do you think is the most unlikely? Simulating a fly brain? Having "long-term" simulation (learning, memory, neuron reconfiguration) in a fly brain? A frog brain? Interfacing with senses or motor controls? Extrapolating that to a human brain? Scanning the brain you want to simulate? (That bit seems like the hardest to do with today's technology, but it's improving) Running the whole simulation at an acceptable speed? (that's probably the bit where there's the most uncertainty, though I would be surprised if *no* function-preserving optimization could be found, considering how the brain remains roughly functional even when neural behavior is changed by drugs)

The extra step - simulating the human brain so that it acts "exactly" like the original, which would be needed to have "true" uploads (where you can consider the uploaded person to be you) requires more precision than merely functional emulations, but functional emulations are enough to severely start disrupting things.

6:26 AM

nick said...

Emile, I've already answered your questions above to the extent possible: the short answer being that specialized computational techniques that are far more efficient will appear long before and be vastly more competitive than "functional brain simulation", to the extent the latter is even feasible. Any benefits or problems posed by these simulations (outside the hope of saving our own consciousnesses, which is beyond the realm of this discussion) will be posed much earlier and to orders of magnitude greater degree by specialized techniques.

10:41 AM

nazgulnarsil said...

this all makes more sense now, still puzzled by this:
"but the remaining degrees of freedom are still likely to be astronomically higher than those of any human-understandable abstraction layer."

is there any reason to believe this? why should the brain necessarily be all that complex? my own inclination is that many of the brain's functions will turn out to be simpler than we thought.

5:30 AM

nick said...

nazgulnarsil, (a) the astronomical number of combinations in the genetic code for the brain, and on top of that of techniques learned culturally, (b) the analog nature of the brain, (c) the computational efficiency of specialization, which implies the brain is composed of many specialized techniques rather than one or a few generalized ones, and (d) the lack of necessity for evolution to create understandable abstraction layers, as is required when human engineers design things. All confirmed by how little psychologists and neurologists still understand about brains, despite centuries of study.

3:12 PM

Tim May said...

A few quick words about Moore's Law.

It was, of course, just an observation about the doubling period in density (and speed, for a while).

Already by around 1980 it was apparent that MOS speeds could not keep doubling every few years (or whatever the precise period was), as the "1/2 CV-squared" switching energies were dominating the power dissipation. But CMOS, which had existed for a while, was adapted to mainstream production technology and this extended the speed increase as well as the density increase.

By around the late 1990s, it was clear that even CMOS was switching so fast such that the devices were no longer acting (in essence) as CMOS....they were drawing current through the CMOS transistors almost continuously. Somewhere around 5 GHz, the transistors are never really "off."

This didn't stop the lithographic density increases, so speeds were backed off to about 3-3.5 GHz and the focus shifted to more cores, now up to 8 cores in mainstream devices and hundreds of (simpler) cores in graphics and experimental CPUs.

A bigger issue is the cost of manufacturing. For more than 40 years there were new and profitable markets sufficient to fund the N semiconductor companies (N is now about 4 for cutting edge companies, those making sub-32 nanometer geometries) and letting them build the increasingly expensive wafer fabs.

(These went from under $20 million around 1970 to more than $5 billion today. I mean a production fab, capable of supplying something like 5% of the world's consumption of cutting-edge chips. Obvoiusly much smaller fabs can exist in niches.)

The key point is this: Moore's Law is not about some automatic doubling or tripling in speed, capacity, whatever in some time. It's been about WHAT THE CONSUMER OR CORPORATE DEMAND WAS.

When I was at Intel in the 70s and 80s, there were many worries--and many precarious financial times!--when it was not at all clear that a new generation of fabs would lead to good profits or to a cratering of the company. In fact, many good competitors cratered during these precarious times. (Mostek, Monolithic Memories, Electronic Arrays, Fairchild Semiconductor, AMI, Inmos, etc.).

So bear this in mind when you look at Moore's Law graphs and assume that this means that in 2025 every homeowner will have the equivalent of 300 billion transistors in his house. They will only if they have bought enough stuff to fund the intervening years.

(Personally, I doubt it. Game consoles, a recent large consumer of lots of transistors, are now about what they were several years ago: XBOX 360, Playstation 3, etc. None of the console makers has yet hinted at a new generation of consoles wiith, for instance, 10x the capacity of the consoles from 2007. Perhaps because consumers just aren't likely to buy yet another $400 game console. We are seeing this in several other market areas as well.)

There was a version of Moore's Law which said that the cost of wafer fabs doubles every two years.

In several more years, if this holds, then a cutting-edge wafer fab will cost $20 billion. How is this funded? Will consumers pay enough for the output of these factories? Will Wall Street pay for machines which shave a few nanoseconds off a trade?

Before you say "Yes," consider all of the other industries which have turned S-shaped largely because paying more and more was no longer an attractive option. Whither the SSTs? Whither the space rockets? Whither the nuclear energy too cheap to meter?

(Oh, and I haven't even mentioned the fact that critical device sizes are now so small that the inevitable fluctuations in dopant atoms may cause some regions to have 37 boron atoms, other regions to have 71 dopant atoms....this is what a region less than 25 nanometers on a side, in three dimensions looks like. And, no, hypothesized inventions like spintronics are not likely to save the day, any more so than a dozen such inventions in past decades ever took over from silicon.)

--Tim May, Corralitos, California

7:39 PM

Tim May said...

I'll add a few comments about Moore's Law, and AI software, some approaches, etc.

(Gad, what am I getting myself into? Some of you may recall my "Rapture of the Future" essay on Extropians from around 1993 or so. I attended most of Ted Kaehler's nanotechnology discussion groups, 1992-1994, and I met various nanotech/cryonics people around 1987-1995. I authored a long article on crypto anarchy in Vernor Vinge's "True Name" reissue about a decade ago. Vernor, whom I admire greatly, my reasons for being skeptical that the Singularity it arriving 20 years away from, fer instance, 1992. Gee, the advocates said it would be here by now....so where is it? I can't wait forever for the Rapture!)

The problem with the techno-Rapture has always been: where does the funding come from? Unless there's some breakthrough which makes new devices essentially free (the canonical "bread machine nanoreplicator" or whatever Greg Bear was "McGuffin"-izing for "Blood Music" (c. 1985), such extension of Moore's Law just doesn't come for free.

Anyway, I already talked about the costs.

And then there's the software problem.

We've made very, very little progress. So what if VisiCalc now runs 20 million times faster than it did in 1982? (Actually, I doubt it does. I run "Numbers" on my 2008-era Mac and it does not run subjectively faster than what I was running in 1982 on an Apple II. At least not for small data sets.)

The real problem was mentioned by several earlier posters: the software.

Now I'm not too worried about the "how do so many transistors communicate with each other?" problem We used to call this "Rent's Rule" (look it up in the Wikipedia)....it relates how the N transistors inside a device communicate by the square root of N (or third root of N or fourth root of N) number of external pins.

Well, the brain does this. And so do actor-oriented and functional programming approaches. (I'm currently a user of Haskell, esp. GHC. But I don't claim it's the perfect AI language. I used Zetalisp in my last years at Intel, then other languages.)

OK, so where is this going?

-- I doubt we'll see any kind of "singularity" in the sense of an AI designing more AIs which design more AIs, and so on, in the next 30 years.

(I didn't think we'd see them 20 years ago when many of my Bay Area friends thought we'd see them within 20 years. Score one for my side.)

-- the software for any meaningful general AI is grossly, grossly behind the hardware.

-- and yet more advocates of AI, strong AI, Friendly AI, Non-Hostile AI, whatever, are going the "Institue" and "Foundation" bullshit routes of fund-raising, proselytizing, scaring, raising funds, etc.

-- we are very, very, very, very far from any kind of AI Singularity. Yeah, the repercussions are potentially bad. But devoting one's life to these possible repercussoins NOW is probably akin to someone setting up the "Steam Age Institute" in 1800.

All of the talk about "let's get it right!" is mostly pointless.

The best way to potentially shape powerful AI--which I guess is dubbed GAI or Goerzel AI or somesuch--is to be a main contributor in the field. Holding conferences with social advocates does not seem to be too useful.

(I left the Extropians, whom I mainly hung out with because some very smart people were hanging out there from around 1992 to 1995, when it was clearer to me that the message was more about the beliefs in living forever, blah blah, than in interesting ideas.)

I'm even less focused on the distant future than Nick is, if that's possible.

--Tim May, Corralitos, CA

8:27 PM

nick said...

I'm even less focused on the distant future than Nick is, if that's possible.

Oh it's certainly possible. Great comments, Tim. It think it's fair to say that never has an industry met customer demand quite so well as the integrated circuit business.

9:43 PM

Emile said...

specialized computational techniques that are far more efficient will appear long before and be vastly more competitive than "functional brain simulation", to the extent the latter is even feasible. Any benefits or problems posed by these simulations (outside the hope of saving our own consciousnesses, which is beyond the realm of this discussion) will be posed much earlier and to orders of magnitude greater degree by specialized techniques.

I already find the brain emulation scenario pretty scary - it seems to imply most humans going the way of the horse (not being able to compete economically with ems). You seem to be saying that it won't happen, not because brain emulation is too hard to achieve, but because humans will already have gone the way of the horse before that because of "specialized" AI (which also involes the question of what happens when most economic agents don't share human values).

If you think it's likely that specialized AI will lead to significant social disruptions, that may not be very far from the beliefs of some "Singularitarians" (I'm not a big fan of the name myself).

2:22 AM

nick said...

Emile:
humans will already have gone the way of the horse before that because of "specialized" AI

We have to talk about what we know, or at least by strong analogy to what we know, or it's just pointless speculation.

So today, some people are trust fund babies. It doesn't matter if they don't have any marketable skills, because they can make a living off the capital gains and dividends thrown off by the humans and machines of the companies they own.

After a very long evolution of hyperspecialization in machines, they will be able to do essentially anything of value that a human would want, excepting those tasks that for emotional reasons require another human (e.g. sports, nail salons). At that point, every human can be a trust fund baby, except people who don't own significant stocks will need to polish the trust funders' nails, or entertain them in sports, or the like, until they've saved up enough money to buy enough stock to live at the standard of living they wish. Of course, people will still have a wide variety of hobbies of varying degrees of intellectuality, usually assisted by myriads of

hyperspecialized machines.

By sharp contrast, horses went the way of the horse because, when IC engines rendered most of them useless, they didn't own anything. Instead the humans who owned them chose to feed most of them to their dogs.

Since there will be nothing like AGI, just quadrillions of different kinds of hyperspecialized machines, it's quite reasonable to say that no ethical issue arises out of humans retaining ownership over their machines and the machines owning nothing.

The wild card, of course, is that I'm assuming that ownership still works, i.e. I'm making the security assumption that humans will still control their machines. Unfortunately, we do have some machines even today that owners lose control over due to poor security. And that's the real issue.

So I submit the only useful questions we can ask are not about AGI, "goals", and other such anthropomorphic, infeasible, irrelevant, and/or hopelessly vague ideas. We can only usefully ask computer security questions. For example some researchers I know believe we can achieve virus-safe computing. If we can achieve security against malware as strong as we can achieve for symmetric key cryptography, then it doesn't matter how smart the software is or what goals it has: if one-way functions exist no computational entity, classical or quantum, can crack symmetric key crypto based on said functions. And if NP-hard public key crypto exists, similarly for public key crypto. These and other security issues, and in particular the security of property rights, are the only real issues here and the rest is BS.

5:51 AM

gwern said...

> When one knows some useful things a priori about particular environments, i.e. when intelligence is specialized, computations that learn things about or solve problems within that environment can often be exponentially or doubly-exponentially more efficient.

Nick, I'm not sure I understand your claims about AIXI.

You say it will be outperformed by specialist tools, but doesn't AIXI specialize itself as it acquires observations and can rule out ever more environments? If we wrote an AIXI, we would have an awful lot of observations to feed into it... (The analogy would be to partial evaluation - AIXI starts off completely general and extremely inefficient, but as inputs are fixed and become static, it becomes faster and specialized.)

And your point about exponentials seems true enough, but IIRC, that exponentiality is for the Kolmogorov complexity of AIXI's environment/the universe, which may be extremely small and doesn't change. So I'm not sure that's a knock-down argument either.

2:14 PM

nick said...

Gwern, we are starting with a world where hundreds of thousands of humans (supported by billions more humans and quadrillions of special purpose algorithms) have already invented a myriad of efficient learning and other algorithms optimized for an extreme variety of special conditions. Then you try to introduce a program you call AGI with the most inefficient (nay, uncomputable) searches (or ways of learning). This putative "AGI" has only access, via simulation or sensing, to a miniscule fraction of these special conditons, so to the extent it can discover or learn anything at all that hasn't already been discovered or learned (since its general learning algorithm is so preposterously inefficient compared to the hyperspecialized learning algorithms already long at work), it can only learn about this tiny subset. So it ends up learning about only a tiny subset (because it is such a slow learner) of a tiny subset (what it can see) of the world. If it is useful at all, and the odds are astronomically against it, it becomes at best just another hyperspecialized computation in a world of hyperspecialized comptuations.

In other words it just adds the miniscule bit of useful novelty it has against all odds discovered to the economy of quadrillions of algorithms dispersed across the planet and hard at work on their own hyperspecialized parts of the economy. The AGI is stuck with its

own extremely inferior general algorithm and, if it got astronomically lucky, one or a handful of new special algorithms that it discovered, and whatever small subset of specialized machines it knows how to handle and has been granted the right to control (another miniscule fraction). There's no magic bullet that makes discovery or other learning take off, things are learned a tiny bit at a time and different hyperspecialized computations working in different parts of the economy learn very different and indeed mutually incomprehensible things.

3:49 PM

gwern said...

AIXI doesn't need to be better than every human in existence, singly or collectively. It just needs to be good enough to pay for its consumption of resources. That doesn't seem very hard at all. (I think *AIXI-MC* might be smart enough to handle paying jobs like breaking CAPTCHAs!)

As for your point about tools. All those tools are already being wielded by absurdly inefficient general purpose intelligences - humans. No human understands more than a tiny fraction of what goes on in its tools (computer). No human sees more than a vanishingly small fraction of its environment, nor do they think much about even that. A few trivial equations will easily outperform us on many tasks.

Yet somehow these stupid beings manage to effectively use them, presumably by simplistic methods like running them and going 'Well, that seemed to work'.

Which AIXI can do.

So, why can't AIXI learn to use these tools as well as the humans? And then better?

4:28 PM

nick said...

gwern:
It just needs to be good enough to pay for its consumption of resources. That doesn't seem very hard at all.

Nope, that's the main point, it is hard, it is astronomically hard, it is *exponentially* slower to learn about an environment in a general way than to learn about a specific environment or aspect of an environment in a specific way suited to that aspect. So our AGI loses, it really can't even start the race, in any competition, whether economic or violent, with already existing hyperspecialized algorithms. Any such general algorithm is merely an intellectual curiosity, it doesn't do anything significantly useful or threatening.

*(I think *AIXI-MC* might be smart enough to handle paying jobs like breaking CAPTCHAs!)*

And this is a perfectly good example. An algorithm especially suited to break CAPTCHAs is going to do the task vastly better than applying some very general purpose learning technique.

All those tools are already being wielded by absurdly inefficient general purpose intelligences - humans.

Nope, we humans ourselves are collections of a variety of special-purpose learning techniques. Much more general than the tools to be sure, but far from the mythological AGI. We already start out with many special algorithms grown as our genetic code elaborates into the astronomically complex network of proteins etc. that is our brain, and we learn many more specialized techniques on top of those as we watch our parents and peers, go to school, gain job experience, and so on, increasingly differentiating our skills from those of our fellow humans the more we learn. Tons of information transfer, and quite a bit of efficient specialized learning, but practically none of what a computer scientist would call learning (discovering new rules from raw data) in anything but particular specialized contexts, e.g. a child learning language from examples he hears.

We will be, and in some cases today already are, wielding tools vastly more capable in any particular task than either ourselves our

any putative AGI. And hopefully we'll be doing this in a security environment exponentially difficult or information-theoretically impossible to crack. OTOH if the security is feasible to crack then again malware hyperspecialized at the job of cracking security is going to do the job vastly more effectively and long before anything resembling AGI. The real threat is hyperspecialized malware, not mythological anthropomorphic software.

5:39 PM

nick said...

Errata:
in any particular task

Less parsing ambiguity if you read that as "in *a* particular task."

5:49 PM

Kevembuangga said...

Another different take about why there cannot be any "logic based godlike infallible AGI" and therefore no Singularity.
Because the world is already "bizarre" enough that it cannot be fully modeled rationally.
By Monica Anderson.

10:05 AM

nick said...

A great lecture, thanks! The first half about "bizarre" (roughly speaking, too complex/chaotic/variable to model rationally) is quite good. The Newtonian easy - to - society hard spectrum corresponds to Hayek/Feynman/et. al.'s critique of "social sciences" as what Feynman called "cargo cult science", that is to say mimicking the techniques of physics to try to look impressive when the subject matter is far too "bizarre" in Anderson's terminology to be explained in terms of such simple (low Kolmogorov complexity, low chaos, etc.) models as those used in the vast majority of physics. That's why I put a far higher intellectual value on highly evolved systems of thought (e.g. law, history, and political and economic views based on long intellectual and practical traditions) than on modern "social sciences".

The second half of Anderson's presentation seems a bit hand-wavy. General techniques like evolutionary algorithms also work better in the simple/easy domains like Newtonian mechanics. And rather than "1 millisecond" a generation in evolutionary computation can take anywhere from microseconds to forever to run based on the feasibility of the fitness function, again with simple mechanics being the microsecond and most aspects of large societies forever. "Bizarre" domains are inherently difficult, and only some of them are even possible to crack with any technique, and these can only be feasibly cracked tiny bits at a time with hyperspecialized techniques.

So I'd say "intuition" is largely just a vast collection of hyperspecialized techniques we haven't discovered yet and that operate at a largely or entirely subconscious level in ourselves (and thus are largely immune to introspection as Anderson well notes).

Thus, when Anderson calls for "enforced diversity" as a protection against "Singularity"/"SkyNet" I see diversity as self-enforcing due to the necessarily hyperspecialized nature of the much more feasible techniques for the reasons described above.

One interesting "intuitionist" technique I haven't seen explored nearly enough in AI is analogy/metaphor/simile. See Lackoff and Johnson's brilliant work on tacit (conceptual) metaphor, *Metaphors We Live By* to see how ubiquitous these are in our (largely subconscious) thinking. Through conceptual metaphor we understand the "bizarre" in terms of the simple and concrete based on our embodied experiences in the world, e.g. via the simple topological relationships we have abstracted from our environment as reflected in prepositions. If somebody wants to mine a probably very lucrative but hardly explored territory in AI, conceptual metaphor is a great topic to dive into.

3:09 AM

Monica Anderson said...

I'm glad some of you have discovered my videos. I also recently tarted writing a series of articles about 21st century AI for h+ magazine and the first one discusses exactly the issue of "friendly AI". It expands a bit on the theme at the end of the ANDIAIR video quoted above. See http://hplusmagazine.com/2010/12/15/problem-solved-unfriendly-ai . I'd love to hear any comments you have on it.

12:17 AM

V V said...

@gwern : You seem to be misunderstanding what AIXI is.

AIXI is a mathematical model of a reinforcement learning agent that satisfies certain optimality criteria under some specific assumptions.

An AIXI agent is uncomputable, thus, as far as we know, unphysical: it needs literaly infinte processing power. It can learn anything because there is no tradeoff involved.

You don't "write" an AIXI. AIXI has no notion of efficiency or speed, it already assumes an infinitely fast hardware with infinite memory.

The moment you try to implement a computable approximation of AIXI, you'll have to include some additional constraint, tradeoff or heuristic. Typically, the more you specialise your algorithms, the better the perforances you'll be able to obtain.

7:56 AM

gwern said...

> The moment you try to implement a computable approximation of AIXI, you'll have to include some additional constraint, tradeoff or heuristic. Typically, the more you specialise your algorithms, the better the perforances you'll be able to obtain.

V V, I don't think I misunderstand this point at all, and I don't think it refutes any of what I said.

7:34 AM

Anonymous said...

The term “Singularity” was coined by astrophysicists to refer to the zero-volume, infinitely dense center of a black hole where matter apparently disappears from our universe. It is a singular event that is so profoundly unique because it defies our laws of physics and nothing else like it (that we know of) occurs in the cosmos. We know that matter can change from one state to another, e.g., water, steam, or ice—gas, liquid, solid, or plasma—and mass can change into energy or vice versa; we can burn coal and make heat, ash, and smoke—but nowhere in our universe other than a black hole does matter 100 percent cease to exist—as far as we know.

Ray Kurzweil has written about an anticipated major shift in human history, an event so revolutionary as to make everything that has gone before it relatively insignificant. Many pundits agree that a paradigm shift in human history of a similar ultra-transcendental nature will occur sometime during the middle of this century but not all agree on its nature—or how it will play out. Scientists and nearly all religionists profess that it will not alter their irreconcilable viewpoints but may suggest points on which they can agree.

Futurists believe the Singularity must occur because knowledge is not only growing exponentially, but at an exponential rate so that a veritable explosion in technological progress must occur at some point. They suggest that the tipping point will come by 2020 when quantum computing enables artificial intelligence (AI) to begin improving its own source code faster than humans can. In that Singular Moment, AI will begin outperforming human intelligence and facilitate the determination of the function of all genes and the process of protein splicing. By this theory, it will be possible to transcend biology by reprogramming our bodies to a more youthful age, perfect health, an IQ of 165-plus and a PhD in any specialized human knowledge of our choice. That means that science will be able to perfect human "nature," overcome all human shortcomings and problems—and defy aging and death—literally forever.

You can buy this book now on any of the following websites:

Strategic Book Publishing Rights Agency: http://sbpra.com/HenryMarkant/

Amazon Books: http://www.amazon.com/Coming-Crises-Their-Solutions-ebook/dp/B00A2WZ4CK/ref=sr_1_1?s=books&ie=UTF8&qid=1357573018&sr=1-1&keywords=coming+crisis+henry+markant

Barnes and Noble Books: http://www.barnesandnoble.com/w/coming-crises-and-their-solutions-henry-markant/1113749628?ean=2940015922875

3:14 PM

saifedean said...

For me, what makes an entity intelligent is the ability to act in its own benefit, as it defines it. A human who decides to eat an apple is intelligent, in that he did something that has improved his subjective well-being. A human who is unable to take actions to further his own well-being is not intelligent. Hence, a man who consumes arsenic and kills himself unwillingly is not intelligent.

As such, I find the notion of Artificial Intelligence to be an oxymoron. If something has been artificially created, whatever actions it carries out will be actions that have been programmed into it by its creator, to serve the creator's needs. You can add all sorts of sensors to process real-world signals and act upon them to a device, and it will act upon them, but only to serve the needs of the creator who made it. It cannot have its own will, because it does not have a life of its own.

There has never been a successful attempt to create life from anything but existing life (as far as I know). So we can really only create more intelligence by creating more life. The quest to create intelligence out of machines is doomed, in my opinion, because it is a contradiction in terms. Machines can become fast, strong, and effective in many tasks, but unless they have synthetic life in them created from something other than existing life, then they are not intelligent. Before thinking about creating artificial intelligence, enthusiasts should consider whether it's possible to create artificial life.

10:33 AM

Simon Majou said...

"when quantum computing enables artificial intelligence (AI) to begin improving its own source code faster than humans can"

https://en.wikipedia.org/wiki/Ponzi_scheme

7:52 AM

Anonymous said...

Nick,

Do the advances of Google's Deep Mind project cause you to re-evaluate your position as expressed above on GAI, or not?

R

6:19 AM

Post a Comment

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)

Simple theme. Powered by Blogger.